**Advances and Applications in Mathematical Sciences**
Volume 21, Issue 11, September 2022, Pages 6495-6511


# A BAYESIAN LOGISTIC REGRESSION MODEL TO ANALYSE THE PERFORMANCE METRICS OF AN INVESTMENT ADVISORY

**S. MYTHREYI KOPPUR[1] and Dr. B. SENTHILKUMAR[2]**

Research Scholar
Department of Statistics
Thanthai Periyar Govt Arts and Science
College (Autonomous), Trichy - 23
(Affiliated to Bharathidasan University, Trichy – 24)

Assistant Professor
Department of Statistics
Thanthai Periyar Govt Arts and Science
College (Autonomous), Trichy - 23
(Affiliated to Bharathidasan University, Trichy – 24)
E-mail: senthilkumarbalan@gmail.com

## Abstract

A financial market agent or intermediary who trades in securities known as stockbrokers, whose primary objective is to acquire and sell stocks for their clients. This study aims at estimating the Net result from the dataset received from a stock brokerage firm-based Chennai. However, there is a minimal or nil attempts to bring performance metrics for investment advisories from historical data based on their investment recommendations. This study has investigated the importance of the response variable Net result, while determining the predictor variable associated with it. A generalized linear model approach with logit link function has been studied using Bayesian framework. From the result of this study, it is observed that the advisory makes a good profit in options trading and the profit is quite high in second Quarter. A special interesting feature to note is odds are higher for Net result ending in Profit, in Sell call recommendations. The whole exercise has been carried out using rstanarm for Bayesian logistic regression modelling.

2020 Mathematics Subject Classification: 62F15.

Keywords: investment advisory, logit link, Bayesian Modelling, logistic regression, rstanarm.
[1]Corresponding author; mythreyi.phd@gmail.com
Received January 27, 2022; Accepted April 12, 2022

## 1. Introduction

Stocks are purchased in order to increase our wealth over time. While some of them believe that stocks are a risky investment, multiple studies have shown that investing in the right equities for a long time (five to ten years) may provide inflation-beating results, making them a better option than real estate or gold. People have short-term strategies when it comes to investing in the stock market. While stocks may be extremely volatile over short periods of time, investing in the right stocks can help traders profit rapidly.

The share market, often known as the stock market (or) equity market, is a collective aggregate of buyers and sellers of equities. In other words, a stock exchange is a market where people may buy and sell shares in companies that are listed on the stock exchange of a given country. Individual units of ownership in a firm are called shares. When a company decides to list on the stock exchange, it normally releases a portion of its equity in order to attract public investment.

In India, there are roughly 23 stock exchanges, with the biggest being the Bombay Stock Exchange (BSE) and the National Stock Exchange (NSE). The Indian stock exchange is presently one of the world's most important stock exchanges. The National Stock Exchange (NSE) is the tenth largest stock exchange in the world, with a market capitalization of USD 3.4 trillion (as of August 2021). NSE's success in this decade, particularly after 2012, has been very remarkable.

Stock brokers serve as a link or point of contact between stock exchanges and their customers. A stockbroker is a financial market agent or middleman who deals in securities. Their main goal is to buy and sell stocks on behalf of their clients. Share brokers might operate for themselves or for a brokerage business. The majority of stockbrokers work for a brokerage business and deal with a variety of individual and institutional clients.

Stockbrokers are frequently compensated on a commission basis; however, this varies by job. Stockbrokers are a term used to describe brokerage businesses and broker-dealers.

A stock broker has to perform his primary responsibilities which are,

- Research
- Buying a stock
- Selling a stock
- Marketing.

The approved stock broker's principal responsibility is to help customers in making educated investment decisions. They provide all conceivable assistance to clients in order to help them invest in securities successfully and profitably. They assist clients in completing the last leg of their investment. They use their expertise to make the stock broking process easier for their clients and to help them make confident stock selections. A stock broker may also assist his clients by providing advice and market news on the finest offers, making the investment process easier. They gain a thorough understanding of the client's needs and financial objectives before recommending the best investment options.

The data for this study is received from a reputed stock brokerage firm, based Chennai. This study focuses on one of the most fundamental and straightforward procedures in the generalised linear model. The Bayesian binary logistic regression model is regarded as an effort at modelling the situation at hand. The main goal is to figure out what causes the causal association between a collection of predictors found in the dataset and a response variable.

Logistic regression is a widely used statistical approach for classifying issues and predicting binary answers in a variety of fields. This paradigm has been effectively applied in a variety of disciplines over the last decade, including business, medical (Chang et al., 2018), and social sciences (Wang et al., 2010). Using Bayesian models will aid the researcher in gaining a better grasp of the model's estimated parameters (Joseph, 2016).

In this paper, we look at how to classify Net result (Profit / Loss) using Bayesian logistic regression (BLR) models and Markov Chain Monte Carlo (MCMC) approaches. In a Bayesian framework, the BLR model combines prior knowledge with the LR model (Hassan et al., 2019). This paper shows how to fit a logistic regression in a Bayesian framework using R and R stan arm. The fundamental benefit of utilising a Bayesian model is that it allows you to combine prior knowledge with data, allowing you to incorporate historical information about unknown parameters and create prior distributions to estimate posterior probabilities and learn about the real value of the parameters (Hassan, 2018). In this work, the performance of Bayesian analysis is evaluated in order to comprehend the performance of a stock broker in terms of his Net result (Profit/Loss). Section 2 contains a typical dataset description; section 3 has a summary of the methodologies and models used in the study; section 4 contains specifics of data analysis and interpretation; and section 5 contains the comments and conclusions.

## 2. Data Set and Description

Brokerage firms play a critical role in financial markets by allowing the efficient transfer of funds from those who need it to those who want it, particularly through securitizations. They also provide investment advising to companies looking to invest in financial assets. As a result, they may be found on both sides of the investment-financing process, helping to produce financial data in the financial markets. The rise of information technology has resulted in increasing rivalry, which has had a substantial impact on the performance of brokerage firms, which are critical in the financial services industry. As a result, it has become critical to look back on the brokerage industry's success in recent years and analyse it in light of market changes. This study was developed in response to this need.

The information used in this study comes from a firm that provides live trading calls based on NSE and BSE data. The trader's CUE programme, which is backed up by analytics and visualisation alerts, assists traders in gaining a better knowledge of the stock market in order to increase profits.

When analysing data, the goal is to learn about the variables and how they relate to the variables of interest. This dataset contains 6111 observations for the year 2020. This is a one-year dataset of 23 variables, 13 of which are categorical in type, that runs January 1, 2020, to December 31, 2020.

**Table 1.** Description of the Metric variables.

| S. No | Variable Name | Description | Nature |
|---|---|---|---|
| 1 | Stocks | Stock Name | Character |
| 2 | Stk_OPN_Date | Date of a stock open | Date |
| 3 | Stk_CLS_Date | Date of a stock closed | Date |
| 4 | Call_Amount | Outcome value - P/L in INR | Numeric |
| 5 | Net_Result | Outcome value - P/L after deduction in INR | Numeric |
| 6 | Entry_Rate | Rate when the call is initiated | Numeric |
| 7 | Year | Year of the transaction | Character |
| 8 | St_date | Date of a stock open (Start) | Date |
| 9 | Ed_date | Date of a stock closed (End) | Date |
| 10 | Day_Diff | No of days a stock kept open | Numeric |

**Table 2.** Description of the Categorical variables.

| S. No | Variable Name | Description | Nature | Details |
|---|---|---|---|---|
| 1 | Product_Type | Segment Codes | Factor | 4 levels |
| 2 | Status_symbol | Call Sequence | Factor | 12 levels |
| 3 | Call_Status | Buy / Sell status when the call is initiated | Factor | 2 levels |
| 4 | Call_Result | Outcome status - P or L | Factor | 2 levels |
| 5 | Entry_Rate_C | Classification of Entry_Rate | Factor | 7 levels |
| 6 | Segment | Name of the segment | Factor | 4 levels |
| 7 | Quar | Quarter (Jan - Dec) of the transaction | Factor | 4 levels |
| 8 | Outcome | Classification of Outcome | Factor | 5 levels |
| 9 | Alerts | Whether the call has Alert | Factor | 2 levels |
| 10 | Direct_Tgt | Whether the call yields Tgt2 directly | Factor | 2 levels |
| 11 | Tgt1 | Whether the call has Tgt1 | Factor | 2 levels |

| | | | | |
|---|---|---|---|---|
| 12 | Net_Result | Outcome status - P or L | Factor | 2 levels |
| 13 | Month_Num | Number 1 - 12 corresponding to a month | Factor | 12 levels |

The major objective in logistic regression analysis is,

• To build a logistic regression model to identify the link between the dependent and independent factors.

• To test whether the independent variables have an effect on the dependent variable.

This study aims on sticking to a commonly used approach for handling categorical response variables and to apply the link functions which helps us in understanding the link between the explanatory and response variables. The goal of this research is to discover a link between Net result and a few predictors. The focus is on estimating the relationship between the variables and presenting various features of Bayesian regression modelling.

The variable Net Result or otherwise Net Result, which is considered as a Response variable is metric in nature. It is then treated as a categorical variable having two categories (Profit and Loss). Figure 1 explains the Net result which ends up with 4106 Profit calls and 2005 loss calls. Overall, the advisor has ended up with 67% of profit through call recommendations. It is clear that the investment advisor has made a good profit for the year 2020.

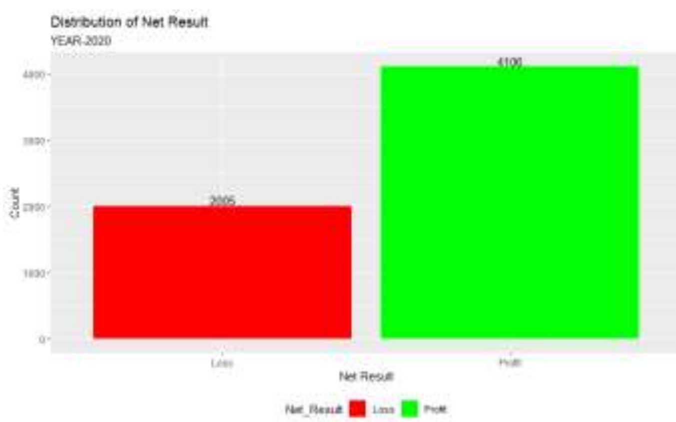


**Figure 1.** Net result for the year 2020.

The predictor or the explanatory variables considered in this study are:

- Call Status
- Call Amount
- Entry_rate_C
- Segment
- Quarter

Call Status is a binary categorical variable having two factors (Buy and Sell). It is the recommendations given by the advisory. There were about 3023 Buy call recommendations and 3088 Sell call recommendations. From Figure 2, it is observed that the overall recommendations given by the advisory ends in profit compared to loss. Especially, compared to Buy call, sell call recommendations results in more profit.

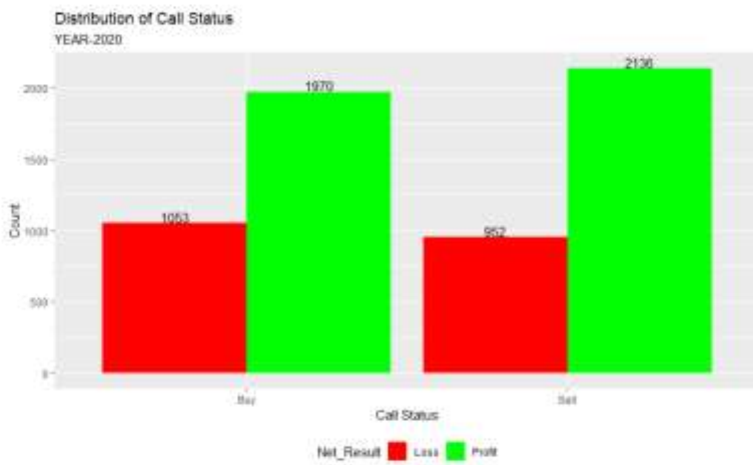


**Figure 2.** Distribution of Call Status with respect to Net Result.

Call amount is a variable which is metric in nature. It represents the outcome value resulting in profit or loss before final calculations. Entry_rate_C is a converted categorical variable from the metric variable which represents a rate at which the call is initiated. It is a categorical variable having seven levels.

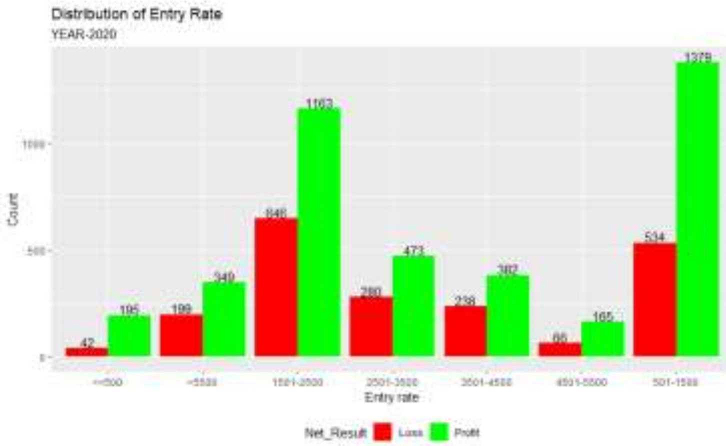


**Figure 3.** Distribution of Entry Rates with respect to Net Result.

From Figure 3, it is observed that in all the entry rate levels, the recommendations ended in profit which is quite higher in the entry rate (1501-2500) and (501-1500).

The variable Segment are securities which represents ownership rights, the advisory deals with four types of segments which are Cash, Futures, Options and Index.

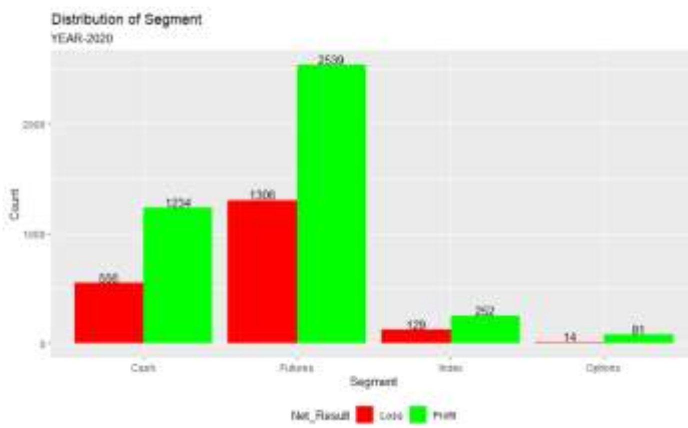


**Figure 4.** Distribution of Segment with respect to Net Result.

From Figure 4, it could be seen that in all the segments the advisor has made a good profit, especially in options trading the reach is high compared to other three segments.

The variable Quarter represents four Quarters (Q1, Q2, Q3, Q4) for the year 2020.

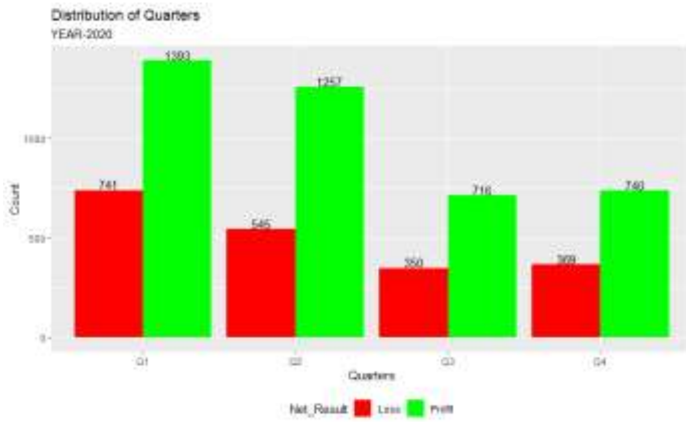


**Figure 5.** Distribution of Quarters with respect to Net Result.

From Figure 5, it could be observed that majority of the recommendations, resulted in profit in all the Quarters throughout 2020, especially higher in Q1 and Q2. The next chapter in detail briefs the important aspects in measures of association and the statistical principle involved.

## 3. Methods and Materials

A regression model with the logistic transformation is called the logistic regression model. Simple logistic regression model can be expressed as follows:

$$\log it(\pi) = \alpha + \beta x$$

$\pi$ is the probability when the outcome variable equals 1. $\text{logit}(\pi)$ is the logistic transformation of the probability of success. Here, $\alpha$ is the intercept and $\beta$ is the logit regression coefficient. Here, instead of directly estimating the dependent variable, we estimate the logistic transformation of the probability of a success also known as the "logarithm of the odds" or "log odds". Odds can be defined as the ratio of two probabilities, the probability of success to probability of failure.

$$Odds = \frac{p}{1-p}$$

In simple logistic regression, we estimate the relationship between an independent variable and the binary response variable on a scale of the logit or log odds.

$$\ln\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta x$$

In multiple logistic regression,

$$\ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_2 x_2 + \ldots + \beta_p x_p$$

where $x_1, x_2, \ldots, x_p$ are the predictor variables and $\beta_1, \beta_2, \ldots, \beta_p$ are the logit coefficients of the predictors.

The nature of the response variable, managing the mean of the response variable, and continue with the linear form are all obvious ways of expanding the idea of this linear model. As a result, the generalized linear model is an extension of the standard linear regression model.

A generalized linear model uses three components:

• Random component, which explains the response variable and its probability distribution.

• Systematic component otherwise called as a linear predictor which is a linear equation on the parameter vector $\beta$ and the explanatory variables $X$.

• Link function which is the mean of the response variable $Y$ that relates linear predictor, $\eta(\mu) = X\beta$.

A linear predictor provides a linear combination of parameters and predictors as,

$$\eta(\mu) = \sum_{j=1}^{k} \beta_j X_{ij}$$

Hence exponentiation gives the odds for success of the response variable to that of failure given a set of $k$ predictors, with the inclusion of constant in the model. Adding more than one numeric predictor might result in a variety of models. Such variable's levels might be dichotomous or polychotomous. In comparison to numerical predictors, the interpretation of estimated regression coefficients of such variables will change. One of the levels will serve as a reference or base level. When the levels are compared to the base level, regression coefficients will show the effective changes in the mean of the response.

## 4. Data Analysis and Interpretation

The purpose of this article was to look at possible analytical models for financial advisers. This study discovered that Bayesian modelling has a lot of promise for delivering effective analytical tools when prior assumptions are made correctly. The first step was to obtain a dataset that reflected the broker's investment recommendations for the year 2020 (January to December). The second and most crucial step is to identify the appropriate variables and a method for dealing with them. The next stage is to determine which associated variables, such as response and predictor variables, are relevant.

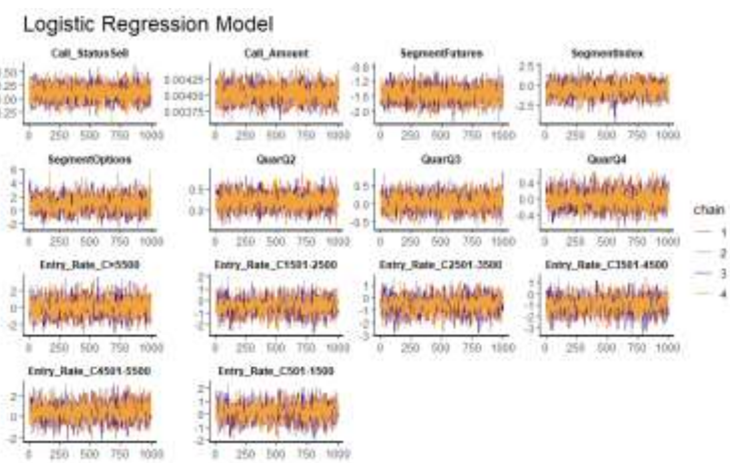


**Figure 6.** Trace plots of the various predictors considered in this study.

It is essential to have a sample that has valid information about the posterior distribution in order to estimate Bayesian models using MCMC. The $R$ packages rstanarm include several critical MCMC diagnostic tests. It is essential to evaluate whether the chains converge to the same place while evaluating convergence. The various colours represent distinct chains, each of which began with a randomly chosen beginning value. A trace plot is the name for this sort of graph. The model comprising the numerous components evaluated in this study does not appear to be divergent, based on the aforementioned trace plot.

The graph does not reveal any alarming departures from normality as a measure of convergence. On the other hand, convergence reveals no significant differences. The posterior sample can be represented using a trace plot, which shows samples over iterations. A trace graphic depicts the sampled data per chain and node throughout iterations. Convergence, on the other hand, indicates that none of the models diverge significantly. The goal is to determine the response variable, which is considered to be Net result in the data set, as well as numerous predictor factors.

**Table 3.** Regression coefficients for the predictors with 97.5% confidence interval of beta under binary logistic regression model with logit link function.

| Binary Logistic regression model with logit link function | | | |
|---|---|---|---|
| Predictors | | 97.5% CI of β | |
| | β | LL | UL |
| Call_Status (Sell) | 1.1151 | 0.8496 | 1.4740 |
| Call_Amount | 1.0040 | 1.0038 | 1.0043 |
| Segment (Futures) | 0.2145 | 0.1463 | 0.3142 |
| Segment (Index) | 0.5874 | 0.0973 | 3.0962 |
| Segment (Options) | 2.3048 | 0.3142 | 22.3807 |
| Quar (Q2) | 1.2503 | 0.8691 | 1.7864 |
| Quar (Q3) | 1.0469 | 0.6786 | 1.6021 |
| Quar (Q4) | 0.9790 | 0.6525 | 1.4729 |
| Entry_Rate_C >5500 | 1.1201 | 0.1708 | 8.4493 |
| Entry_Rate_C1501-2500 | 0.6377 | 0.1817 | 2.1722 |

| | | | |
|---|---|---|---|
| Entry_Rate_C2501-3500 | 0.5803 | 0.1536 | 2.0074 |
| Entry_Rate_C3501-4500 | 0.3803 | 0.0890 | 1.5169 |
| Entry_Rate_C4501-5500 | 1.6010 | 0.3215 | 7.6994 |
| Entry_Rate_C501-1500 | 1.2035 | 0.3282 | 4.0539 |

The binary logistic regression model has been fitted for the data set and the estimates of the model parameters are presented in Table 3. It can be observed from the Table 3 for the logit link function that the predictors used in the study are call status, call amount, segment, Quarter and Entry rate. The odds ratio is greater than 1 for the predictors call status and call amount. For the predictor segment, a categorical predictor which has four levels in which cash is considered to be the base level. Considering the other three levels, options under segment has a greater odds ratio of about 2.3048. While considering the predictor Quarter, 4th Quarter has an odds ratio estimate less than 1. Considering the predictor entry rate, except for the entry rates 1501-2500, 2501-3500 and 3501-4500 all the other entry rates have an odds ratio value greater than 1.

Now in detail the regression coefficients could be interpreted as follows:

- The predictor Call Status having two levels (Buy/Sell), in which call status (Buy) is considered as a reference level. The odds ratio estimate for call status (Sell) is 1.1151. While considering the sell call recommendations (compared to Buy call recommendations) it is associated with a 11.5% increase in the odds of Net result, resulting in profit.

- Considering the predictor Call Amount which is a metric variable, the odds ratio estimate is 1.0040. For a one-point increase in the Call Amount, it is observed that there is a 0.40% increase in the odds of Net result, resulting in Profit.

- Considering the predictor Segment, in which Cash is considered as a reference level, the odds ratio estimate for Futures is 0.2145. Given the recommendations under Futures (compared to cash) is associated with a 79% decrease in the odds of Net result, resulting in Profit.

The odds ratio estimate for Index is 0.5874, given the recommendations for Index in the year 2020 (compared to Cash) is associated with a 41% decrease in the odds of Net result, resulting in Profit.

The odds ratio estimate for Options is 2.3048, given the recommendation for Options (compared to Cash) is associated with a 130% increase in the odds of Net result, resulting in Profit which is a quite interesting feature to note.

• Considering the predictor Quarter, in which the 1st Quarter is considered as the base level. The odds ratio estimate for the 2nd Quarter is 1.2503. Being in the second Quarter (as compared to 1st Quarter) is associated with a 25% increase in the odds of Net result, resulting profit.

The odds ratio estimate for 3rd Quarter is 1.0469, being in the 3rd Quarter (as compared to 1st Quarter) is associated with a 5% increase in the odds of Net result, resulting in Profit.

The odds ratio estimate for 4th Quarter is 0.9790, being in the 4th Quarter (as compared to 1st Quarter) is associated with a 2% decrease in the odds of Net result, resulting in Profit.

• Considering the predictor Entry rate which was metric in nature in nature and treated in to categorical variable having seven levels in which <=500 is considered as a reference level. The odds ratio estimate for the entry rate (501-1500) is 1.2035 indicating that, when compared to the base level is associated with a 20% increase in the odds of Net result, resulting in Profit.

The odds ratio estimates for the entry rate (1501-2500) is 0.6377 indicating that, when compared to base level is associated with a 36% decrease in the odds of Net result, resulting in Profit.

The odds ratio estimates for the entry rate (2501-3500) is 0.5803 indicating that, when compared to base level is associated with a 42% decrease in the odds of Net result, resulting in Profit.

The odds ratio estimate for the entry rate (3501-4500) is 0.3803 indicating that, when compared to base level is associated with a 62% decrease in the odds of Net result, resulting in Profit.

The odds ratio estimate for the entry rate (4501-5500) is 1.6010 indicating that, when compared to base level is associated with a 60% increase in the odds of Net result, resulting in Profit which is quite notable feature compared to other entry rates.

The odds ratio estimate for the entry rate (>5500) is 1.1201 indicating that, when compared to base level is associated with a 12% increase in the odds of Net result, resulting in Profit.

## 5. Conclusion

This study has adopted Bayesian methods for Binary Logistic regression model. The major aim lies in understanding the Net Result for the year 2020 resulting in Profit/Loss which depends on predictor variables such as Call status, Call Amount, Entry rate, Segment and Quarters. The whole exercise has been carried out using R stan and R stan arm and the interpretation of the coefficients have been discussed in the previous chapter. The interesting notable feature to understand the odds of Net result, resulting in profit are as follows:

- Sell Call recommendations results in yielding profit when compared to Buy Call recommendations.

- The predictor variable Segment having four levels Cash, Futures, Options and Index. This study clearly shows that the advisor makes a good profit in Options trading compared to the other three categories of trading, which is a special important feature to note.

- When considering the Quarters for the year 2020, the advisory makes a good profit in 2nd Quarter compared to the other three Quarters.

- In the Entry rate levels, the special feature to note is in the Entry rate Category (4501-5500) is associated with a 60% increase for the Net Result to be profit.

Overall while considering the above points, it is worth mentioning the special features of the investment advisory. From this study it is clear that when the dependent variable Net result is considered, there are many predictor variables which influence the response variables and it is very clear to understand the reasons behind the Net result, resulting in profit.

## References


[1] J. D. Stoffels, Stock recommendations by investment advisory services: Immediate effects on market price, Financial Analysts Journal 22(2) (1966), 77-86.

[2] J. E. Bodenman, Firm characteristics and location: The case of the institutional investment advisory industry in the United States, 1983-1996, Papers in Regional Science 79(1) (2000), 33-56.

[3] S. Vraneš, M. Stanojević, V. Stevanović and M. Lučin, INVEX: investment advisory expert system, Expert Systems 13(2) (1996), 105-119.

[4] A. Jaretzki, The Investment Company Act: Problems Relating to Investment Advisory Contracts, Virginia Law Review (1959), 1023-1037.

[5] J. E. Bodenman, The suburbanization of the institutional investment advisory industry: Metropolitan Philadelphia, 1983-1993, The Professional Geographer 50(1) (1998), 112-126.

[6] S. M. Tinic and R. R. West, The securities industry under negotiated brokerage commissions: Changes in the structure and performance of New York Stock Exchange member firms, The Bell Journal of Economics (1980), 29-41.

[7] C. T. B. Ho, and K. B. Oh, Measuring online stock broking performance, Industrial Management and Data Systems 108(7) (2008).

[8] H. F. Gholipour and M. N. Razali, Determinants of financial performance of real estate brokerage industry in Iran, International Journal of Housing Markets and Analysis, (2017).

[9] J. Ghosh, Y. Li and R. Mitra, On the use of Cauchy prior distributions for Bayesian logistic regression, Bayesian Analysis 13(2) (2018), 359-383.

[10] W. DuMouchel, Multivariate Bayesian logistic regression for analysis of clinical study safety issues, Statistical Science 27(3) (2012), 319-339.

[11] H. D. G. Acquah, Bayesian logistic regression modelling via markov chain monte carlo algorithm, Journal of Social and Development Sciences 4(4) (2013), 193-197.

[12] C. Chen, G. Zhang, X. C. Liu, Y. Ci, H. Huang, J. Ma, and H. Guan, Driver injury severity outcome analysis in rural interstate highway crashes: A two-level Bayesian logistic regression interpretation, Accident Analysis and Prevention 97 (2016), 69-78.

[13] B. Nandram, L. Chen, S. Fu and B. Manandhar, Bayesian logistic regression for small areas with numerous households, arXiv preprint (2018), arXiv:1806.00446.

[14] K. Vaitheeswaran, M. Subbiah, R. Ramakrishnan and T. Kannan, A comparison of ordinal logistic regression models using Classical and Bayesian approaches in an analysis of factors associated with diabetic retinopathy, Journal of Applied Statistics 43(12) (2016), 2254-2260.

[15] Q. M. Abdulqader, Applying the binary logistic regression analysis on the medical data Science Journal of University of Zakho 5(4) (2017), 330-334.

[16] E. J. Benjamin, S. S. Virani, C. W. Callaway, A. M. Chamberlain, A. R. Chang, S. Cheng and P. Muntner, Heart disease and stroke statistics 2018 update: a report from the American Heart Association, Circulation 137(12) (2018), e67-e492.

[17] J. Wang, S. Kumar and S. F. Chang, Sequential projection learning for hashing with compact codes, (2010).

[18] S. Heydari, L. Fu, L. Joseph and L. F. Miranda-Moreno, Bayesian nonparametric modeling in transportation safety studies: applications in univariate and multivariate settings, Analytic methods in accident research 12 (2016), 18-34.